Speech Technology For Supporting Community-Based Endangered Language Documentation

Robbie Jimerson[1,2], Richard Hatcher[3], Raymond Ptucha[2], Emily Prud'hommeaux[2,4]

[1]Seneca Nation of Indians, [2]Rochester Institute of Technology

[3]State University of New York at Buffalo, [4]Boston College

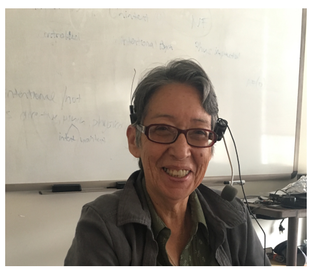
PROJECT GOALS

- Explore **automatic speech recognition (ASR)** for supporting transcription of **Seneca**
- Evaluate the **speed and accuracy** of ASR-supported transcription as a function of language skill.
- Collect data on **individual preferences** for correcting ASR output vs. producing transcripts "from scratch".

ONÖDOWA'GA:' GAWËNÖ'

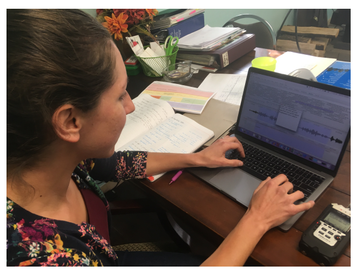
- **Seneca**: member of the Iroquoian language family.
- **~50** first-language speakers, **~100+** learners.
- Spoken primarily in **Western NYS and Ontario**.
- **Polysynthetic** morphology, highly agglutinative.
- **Complex morphophonology**.

TRANSCRIPTION PIPELINES



Sandy Dowdy
Seneca elder

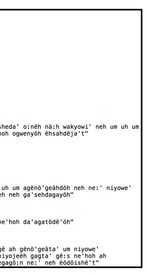
KALDI
+ in-house scripts



Transcription of raw audio with ELAN

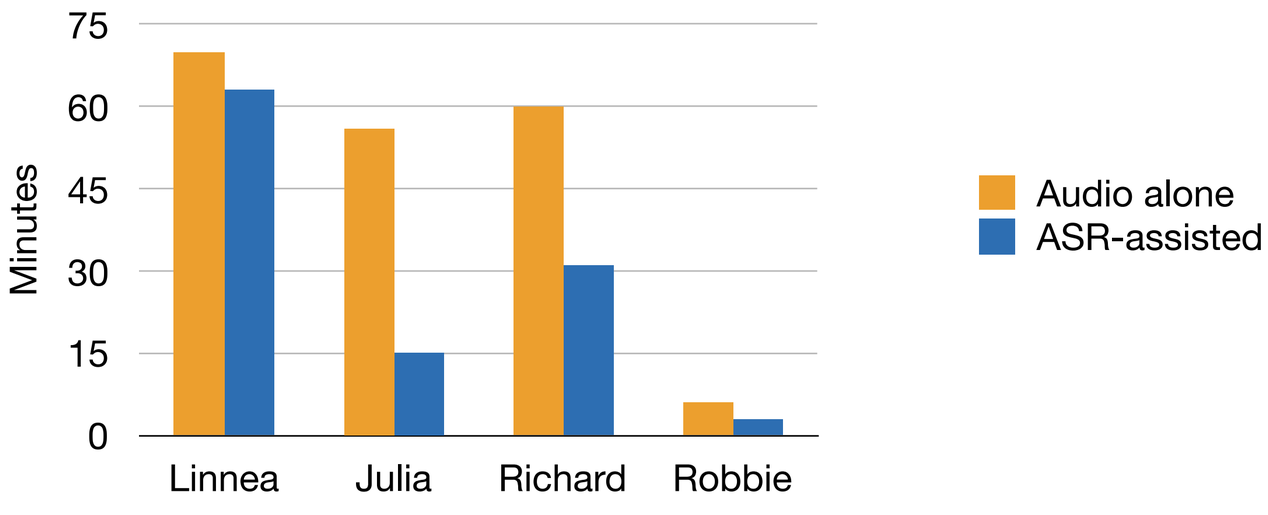
TextGrid of ASR output

ASR-assisted transcription in Praat

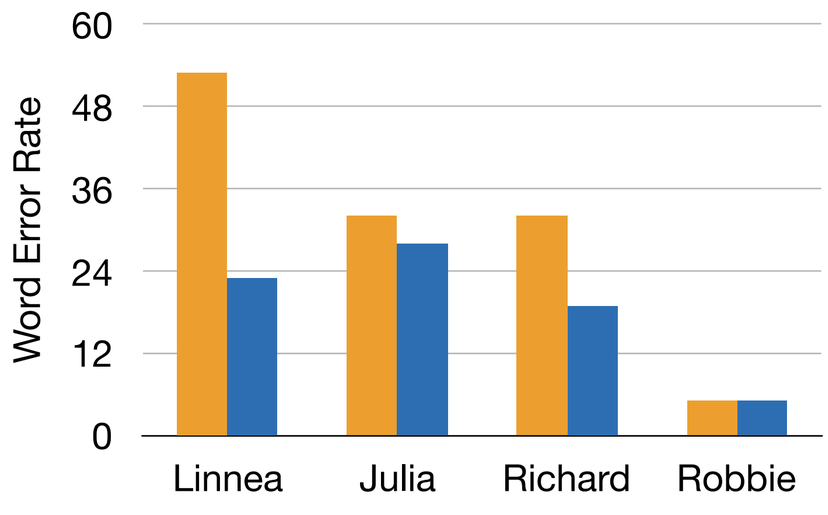


Final TextGrid

USABILITY STUDY: ASR-ASSISTED TRANSCRIPTION



Time required to transcribe in minutes. Transcription with ASR is substantially faster for most participants



Word error rate (WER). WER is substantially lower when using ASR for most participants.

Linnea	undergrad linguistics RA	"I had a difficult time with the ASR, because I spent more time cross-checking the transcription than actually just transcribing."
Julia	undergrad linguistics RA	"Using ASR, I was able to focus on comparing the audio to the transcription rather than trying to perceive what was being said."
Richard	PhD student researching Cayuga	"I preferred the ASR because it took care of some of the necessary steps, e.g., segmenting speech into utterance units."
Robbie	fluent Seneca L2 speaker	"The only thing that slowed me down using ASR was having to copy and paste around the incorrect words."

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1761562.

We are grateful for the support and generosity of the elders of the Seneca Nation of Indians.